To be Published in IEEE Intelligent Systems, Special Issue on e-Science

"Semantic Services" for Grid Based, Large-Scale Science

William E. Johnston[a]
Lawrence Berkeley National Laboratory and
NASA Ames Research Center

Keywords: Grid services, Semantic Grid, science application requirements

Abstract

This paper briefly describes Grid technology for supporting large scale science. It examines several examples of how the process of science must evolve over the next five to ten years in order to facilitate the next steps in scientific discovery. In this context it examines the need and role of semantic description, management, and manipulation of science simulations and data. In conclusion it provides several examples of the potential (even essential) value of semantic tools in dealing with the greatly increased complexity of the multi-disciplinary simulation and data environments required for next generation science.

1 Introduction

Grid technology [1, 2] has evolved over the past several years and is merging with Web Services to provide the mechanisms and infrastructure needed for a standardized and componentized approach to building the large-scale, distributed "virtual" systems and organizations that are necessary for large-scale science. This software technology will knit the hardware, data, and resources into an infrastructure that substantially simplifies building science applications and supporting scientific collaborations that include large scale computing systems, data archives, and instruments that span many different projects, institutions, and countries. This infrastructure will be in the form of Grid based services that can be integrated with the user's work environment, and that enable uniform and highly capable access to these widely distributed resources. These services will integrate transient-use resources like computing systems, scientific instruments, and data caches (e.g., as they are needed to perform a simulation or analyze data from an experiment); persistent-use resources, such as databases, data catalogues, and archives, and; collaborators, whose involvement will continue for the lifetime of a project or longer.

However, as we begin to understand and deploy the capabilities of the Grid Services environment, and how the science community is going to use it, we are also seeing that further capabilities are needed for an infrastructure that will enable the next generation of the process of science.

Beyond the basic Grid Services, and the many application oriented services that are being built in this environment, other "basic" functionality is needed. This includes, e.g., virtual data services [3], application composition frameworks such as the Common Component Architecture[b] XCAT [4] that manage several styles of connectivity between componentized services, Web based portal builders such as XPortlets [5] that provide componentized tools for building graphical user interfaces (GUIs) in the Web and Grid environment, etc.

---

[a] wejohnston@lbl.gov, http://dsd.lbl.gov/~wej
USMail: Lawrence Berkeley National Laboratory, MS 50B-2239 , Berkeley, CA, 94720, USA

[b] The Common Component Architecture is a DOE funded program to define a standard set of behaviors and interface conventions so that a high-performance component framework can be built that allows composing components to build a running application. The XCAT framework uses a Grid Services approach to extend CCA to the distributed computing environment.

However even with all of this, we are still left with a big problem in making all of these tools usable in the science environment. We must provide mechanisms to structure and manipulate various representations of the wealth of available application services, tool services, and data. A promising approach is that of the Semantic Web, where the AI community's work in producing and manipulating discipline oriented descriptions of the semantic aspects of services and data, are being applied to XML based descriptions of applications components (Web Services) and data. We envision tools based on this work that provide for automatic checking of the validity of sequences of composed operations and data, the automatic construction of intermediate steps in a loosely specified sequence, and the automatic construction of sequences of operations that are consistent with a discipline model that represents the permitted relationships among simulation and analysis operations and data for particular disciplines such as climatology of high energy physics. We will refer to these tools as semantic services.

This paper characterizes some aspects and evolution of the science milieu that drives the need for all of these technologies, and provides some requirements for semantic services. Examining two examples of today's process of science and how these science communities anticipate those processes evolving in order to facilitate the next generation scientific discovery illustrates some of the requirements. At the end of the paper a description and analysis of several scenarios illustrates both the need and the potential of integrating semantic tools with the currently evolving Grid environment.

Discipline Portals / Frameworks (problem expression; user state management; collaboration services; workflow engines; fault management)
Applications and Utility Services (domain specific and general components)
Language Specific APIs (Python, Perl, C, C++, Java)
Grid Collective Services / Open Grid Services Architecture (resource brokering; resource co-allocation; data cataloguing, publishing, subscribing, and location management; collective I/O, job management, workflow engines, and component managers)
Core Grid Functions / Open Grid Services Interface (resource discovery; resource access; authentication and security; event publish and subscribe; monitoring / events)
Security Services
Communication Services
Hosting Environments (the local services that support the style of process initiation and management on remote systems – e.g. the C shell and OGSI)
Resource Managers (export resource capabilities to the Grid, handle execution environment establishment, hosting, etc., for compute resources)
Physical Resources (computers, data storage systems, scientific instruments, communication networks, etc.)

Figure 1. The Elements of Grids

2 Grid Technology

The overall motivation for current large-scale, multi-institutional Grid projects is to enable the resource and human interactions that facilitate large-scale science and engineering such as aerospace systems design, high energy physics data analysis [6], climatology, large-scale remote instrument operation [7], collaborative astrophysics based on virtual observatories [8], etc. In this context, the goal of Grids is to provide significant new capabilities to scientists and engineers by facilitating routine construction of information and collaboration based problem solving environments that are built on-demand from large pools of resources.

Functionally, Grids provide tools, middleware, and services for:

- building the application frameworks that allow discipline scientists to express and manage the simulation, analysis, and data management aspects of overall problem solving

- providing a uniform look and feel to a wide variety of distributed computing and data resources
- supporting construction, management, and use of widely distributed application systems
- facilitating human collaboration through common security services, and resource and data sharing
- providing remote access to, and operation of, scientific and engineering instrumentation systems
- managing and securing this computing and data infrastructure as a persistent service

This is accomplished through two aspects: 1) A set of uniform software services that manage and provide access to heterogeneous, distributed resources, and, 2) a widely deployed infrastructure. The layered architecture of Grids is depicted in Figure 1.

The international group working on defining and standardizing Grid middleware is the Global Grid Forum ("GGF," [9]) that now consists of some 700 people from some 130 academic, scientific, and commercial organizations in about 30 countries. GGF involves both scientific and commercial computing interests

3 Science Case Studies

The US Dept. of Energy's Office of Science[a] recently undertook to characterize how the process of doing the sort of large scale science that is DOE's mission must change in order to support advances in that science. In a workshop in August, 2002 [10] the issues were analyzed and requirements set out for networking and middleware. Subsequently, two more workshops [11] [12] proposed approaches to meet these requirements. A fourth workshop examined the computing requirements and approach [13].

This section presents the requirements analysis from two of the case studies in the workshops as they relate to semantic services.

Climate Modeling Requirements[b]

To better understand climate change, we need better climate models providing higher resolution and incorporating more of the physical complexity of the real world. Over the next five years, climate models will see a great increase in complexity, for example in work such as the North American Carbon Project (NACP), which endeavors to fully simulate the terrestrial carbon cycle.

These advances are driven by the need to determine future climate at both local and regional scales as well as changes in climate extremes—droughts, floods, severe storm events, and other phenomena. Over the next five years, climate models will also incorporate the vastly increased volume of observational data now available (and even more in the future), both for hind casting and intercomparison purposes. The result is that instead of tens of terabytes of data per model instantiation, hundreds of terabytes to a few petabytes (10^{15} bytes) of data will be stored at multiple computing sites, to be analyzed by climate scientists worldwide. Middleware systems like the Earth System Grid [14], and its descendents, must be fully utilized in order access and manage such large, distributed, and complex pools of observational and simulation data.

---

[a] The Office of Science supports basic research in all fields of science, with emphasis in the physical, life, and environmental sciences, at the DOE National Laboratories and at Universities. Its annual budget is about $US 3.5B. http://www.er.doe.gov/

[b] This section is based on material from Gary Strand (strandwg@ucar.edu), National Center for Atmospheric Research, and was adapted from "High Performance Network Planning Workshop," 2002, DOE Office of Science, and on material from Tim Killeen of NCAR, presented at the middleware workshops "Blueprint for Future Science Middleware and Grid Research and Infrastructure," 2003, LSN-MAGIC.

In the period five to ten years out, climate models will again increase in resolution, and many more components will be integrated. Climate models will be used to drive regional-scale climate and weather models, which require resolutions in the tens to hundreds of meters range, instead of the hundreds of kilometers resolution of today's Community Climate System Model (CCSM) and Parallel Climate Model (PCM).

Better climate modeling requires that the many institutions working on various aspects of the climate be able to easily describe, catalogue, and seamlessly share the knowledge and the vast amounts of data that underlay the knowledge in order to facilitate the required interdisciplinary collaboration. Further, all of the sub-models must interoperate in ways that represent how the elements that make up the climate interact.

As climate models become more multidisciplinary, scientists from oceanography, the atmospheric sciences, and other fields, will collaborate on the development and examination of more realistic climate models. Biologists, hydrologists, economists, and others will assist in the creation of additional components that represent important but as-yet poorly understood influences on climate that must be coupled with climate models as illustrated in Figure 2.

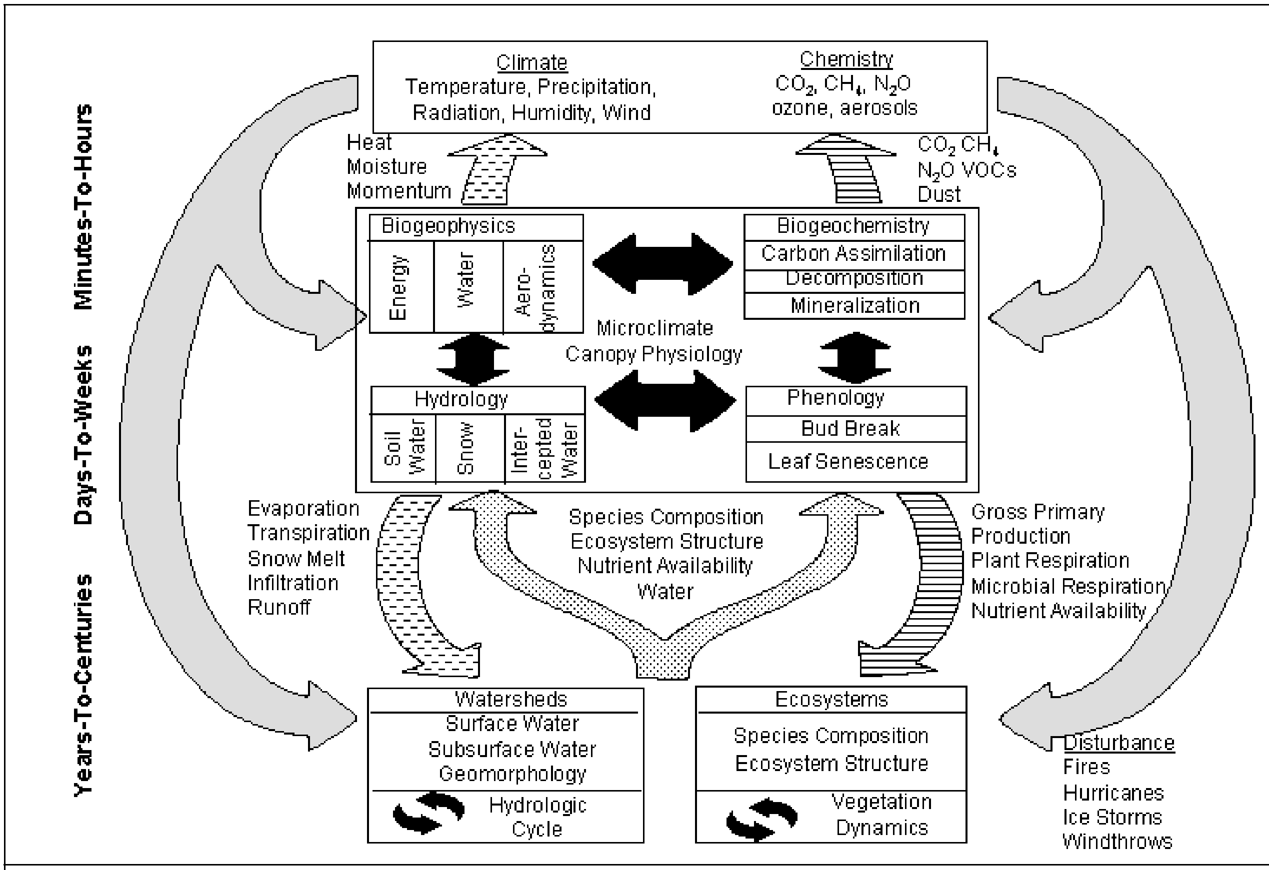

Figure 2. The Complexity of a "Complete" Approach to Climate Modeling Involves the Many Interacting Processes and Data of Terrestrial Biogeoscience
(Courtesy Gordon Bonan, NCAR: *Ecological Climatology: Concepts and Applications*. Cambridge University Press, Cambridge, 2002.)

There will be a true carbon cycle component, where models of biological processes will be used, for example, to simulate marine biochemistry and fully dynamic vegetation. These scenarios will include human population change, growth, and econometric models to simulate the potential changes in natural resource usage and efficiency. Additionally, models representing solar processes will be integrated to better simulate the incoming solar radiation.

The many specialized scientific groups that work on the different components that go into a comprehensive, multi-disciplinary model, build specialized software and data environments that will almost certainly never all be homogenized and combined on a single computing system. Almost all such multidisciplinary simulation is inherently distributed, with the overall simulation consisting of software and data on many different systems combined into a virtual system by using tools and facilities for building distributed systems.

This future process of science is enabled by a set of capabilities that result from combining various computing, communication, data storage systems, with Grid services:

- Computing capacity adequate for a task is provided at the time the task is needed by the science – in particular, supercomputers must be able to be incorporated into "virtual" systems – so that the simulations whose components run on supercomputers may integrate with the many different computing systems of the science community,
- Data capacity sufficient for the science task is provided independent of location, and managed by information systems needed for building, maintaining knowledge bases, and sharing them among disciplines,
- Accommodating the fundamentally distributed nature of the science community with remote access to computing and data and distributed collaboration tools,
- Communication capacity and capability sufficient to support the aforementioned transparently to both systems and users
- Virtual data catalogues and work planners for automatically reconstituting derived data on demand
- Software services providing a rich environment that give scientists the ability to build the multi-disciplinary simulations in ways that are natural to the scientific process, rather than having to focus on the details of managing the underlying computing, data, and communication resources.

In addition to these capabilities, the vision of the future process of science in the climate community requires having the informed interoperation of diverse sub-models, and integration of the knowledge of many disciplines, so that a realistic overall model of climate can make predictions that have great value to human society. Constructing and managing the multi-disciplinary models needed to accomplish this will require tools that can use sub-discipline knowledge to assist in structuring the multi-component processing needed for comprehensive simulations. That is, there need to be tools that can not only build and manipulate complex domain models, but that can also guide the interactions of many different domain models. This sort of semantic service, that addresses building and managing models whose components are themselves complex models, is called category 3 in the discussion at the end of this paper.

The complexity of climate is typical of most macro scale phenomenon from cosmology to cellular function, and so the issues raised by climate modeling, when looking at how the process of science must evolve, are characteristic of much of science.

High-Energy Physics Requirements[a]

The major high-energy and nuclear physics (HENP) experiments of the next twenty years will break new ground in our understanding of the fundamental interactions, structures, and symmetries that govern the nature of matter and space-time. The largest collaborations today – such as the CMS and ATLAS collaborations – are building experiments[b] for CERN's Large Hadron Collider (LHC) and encompass 2000 physicists from 150 institutions in more than 30 countries.

High-energy and nuclear physics problems are the most data-intensive known. The current generation of operational high energy physics experiments at SLAC (BaBar) and FermiLab (D0 and CDF), as well as the nuclear physics experiments at the Relativistic Heavy Ion Collider (RHIC) program at Brookhaven National Laboratory, face many data and collaboration challenges. BaBar, for example, already has accumulated datasets approaching a petabyte. These datasets will increase in size by a factor of a thousand within the next decade. Hundreds to thousands of scientist-developers around the world continually develop software to better select candidate physics signals from the detector data, better calibrate the detector, and better reconstruct the quantities of interest. The globally distributed ensemble of computing and data facilities available to HENP, while large by any standard, is less than the physicists require to do work in a fully creative way. There is thus a need to solve the problem of managing global resources in an optimal way to maximize the potential of the major experiments for breakthrough discoveries.

Collaborations on this global scale would not have been attempted if the physicists could not plan on highly capable networks to interconnect the physics groups throughout the life cycle of the experiment and to make possible the construction of Grid middleware with data intensive services capable of providing access, processing, and analysis of massive datasets. The physicists also must be able to count on highly capable middleware to facilitate the management of worldwide computing and data resources that must all be brought to bear on the data analysis problem of high-energy physics.

To meet the technical goals, priorities have to be set, the system has to be managed and monitored globally end-to-end, and a new mode of "human-Grid" interactions has to be developed and deployed so that the physicists, as well as the Grid system itself, can learn to operate optimally to maximize the workflow through the system. Developing an effective set of trade-offs between high levels of resource utilization and rapid turnaround time, plus matching resource usage profiles to the policy of each scientific collaboration over the long term, present new challenges (new in scale and complexity) for distributed systems.

This will involve

- managing authorization to access secured, worldwide resources
- data migration in response to usage patterns and network performance
- naming and location transparency for data and compute resources
- direct network access to data management systems
- publish / subscribe and global discovery
- monitoring to enable optimized use of network, compute, and storage resources

---

[a] This section is based on material by Julian J. Bunn (julian@cacr.caltech.edu), Center for Advanced Computing Research California Institute of Technology, and Harvey B. Newman (newman@hep.caltech.edu), Physics, California Institute of Technology, and was adapted from "High Performance Network Planning Workshop," 2002, DOE Office of Science.

[b] HENP experiments typically consist of an atomic particle accelerator and a detector specialized to measuring certain properties of elementary particles as they emerge from some sort of a collision: beam-beam, beam-target, etc. The experiment name (e.g. "CMS") indicates both what accelerator is involved (e.g. the Large Hadron Collider at CERN), and what detector is involved (e.g. the Compact Muon Solenoidal detector on the LHC). To find out more about the experiments mentioned here just do a Web search on the experiment name.

- policy based scheduling / brokering for the ensemble of resources needed for a task
- automated planning and prediction to minimized time to complete tasks that includes track world-wide resource usage patterns to maximize utilization

In the context of semantic services, the planning for optimal utilization of resources is increasingly (and of necessity) being addressed through the use of AI based planning techniques as described, e.g., in [15] and [16].

However, at the highest level of problem solving abstraction, where the physicists interact with data that is as highly refined as possible using automated techniques, there remains a need to provide what the science community typically refers to as knowledge management. Consider the following example[a]:

> *"HEP experiments collect specific types of data for the particles that result from high energy collisions of the protons, electrons, ions, etc. that are produced by the accelerators. The types of data are a function of the detector and include things like particle charge, mass, energy, 3D trajectory, etc.*
>
> *However much of the science comes from inferring other aspects of the particle interactions by analyzing what can be observed. Many quantities that are derived from what is observed are used in obtaining the scientific results of the experiment. In doing this more abstract analysis, the physicist typically goes through a process like the following.*
>
> *Events of interest are usually characterized by a combination of jets of particles (coming from quark decays) and single particles like electrons and muons. In addition, we look for missing transverse energy (an apparent failure of momentum conservation) that would signal the presence of neutrinos that we cannot detect.*
>
> *The topologies of individual events follow some statistical distribution, so it is really the averages over many events that are of interest. In doing the analysis, we specify what cone angle would characterize a jet, how far one jet needs to be from another (in 3-dimensions), how far from the single particles, how much missing transverse energy, the angles between the missing energy vector and the other particles, etc.*
>
> *What I would like to see is a set of tools to describe these topologies without typing in lots of code, e.g. a graphical interface that lets you draw the average event and trace out how statistical variations would affect that. We do simulation of interesting processes and they guide the selection of events, so we would want to learn from that as well.*
>
> *In order to transform these sorts of queries into combinations of existing tools and appropriate data queries, some sort of knowledge-based framework is needed."*

This sort of semantic service, that organizes operations within a single domain model, is called category 2 in the discussion at the end of this paper.

4 Semantic Services

In order to realize the benefit of a componentized science simulation environment that is rich in discipline data we must provide several capabilities related to automatic query structuring. That is (at least initially), the semantic services noted above are primarily related to the automatic verification and structuring of various forms of queries within the fairly well defined and stylized environment of a scientific discipline.

[a] This example is courtesy of Stewart Loken, Physics Division, Lawrence Berkeley National Lab.

Category 1: The first capability that is needed is to be able to check the validity of complex sequences that are manually constructed by the user, and to provide guidance if they are incorrectly structured.

A scientist may know perfectly well how to formulate an abstract sequence of operations on data that will answer a question or obtain a desired result in terms of the science analysis steps. However, the exact forms of analysis / simulation components, and data that are available, may not be directly suitable for the desired sequence at the science level, or the available components may only produce the desired transformation if invoked in certain ways, etc. The specifics of permitted connections between components or the specific formats of the data need to be encoded in semantic models that can provide higher-level constraints on interrelationships, and informs the user of constraint violations.

We have seen some precursors of this sort of capability in, e.g., graphical model builders that enforce semantic data compatibility for a few data types when building the workflow network that represents the discipline model. That is, the user gets some help in the form of a graphical programming language that enforces certain types of constraints among the building blocks. However, this approach is rigid and very limited in the range of interconnection relationships that can be represented.

A generalization of this is needed that provides detailed descriptions of the data through metadata and XML schema, and corresponding descriptions of the nature of the data needed as the input of a component and produced as output. Given this, tools are needed that can check the validity of constructed networks though a complete compatibility analysis of input, output, and data types and formats as well as the semantic relationships among the components. Incompatibilities should be reported in a meaningful way that indicates what components may correctly interact, what types of data characteristics are needed for a given operation, what data formats are available, etc.

Category 2: At the next level of capability it should be possible to automatically build up relatively simple composite operations from libraries of simpler ones based on the semantic relationships of the components. That is, given the semantic relationships among a fairly limited and well defined set of primitive operations and data within a single, well defined discipline model, semantic tools should permit the automatic construction of compound operations that appropriately transform the data by invoking primitive operations in the correct order to provide the desired result. For example, if a user wants the linear velocity components of a particle, and the available data provides angular momentum and mass then is should be possible to automatically assemble the sequences of transformations that derive linear velocity. This sort of capability would address the requirements of the example query given in the High Energy Physics case study, above.

Category 3: The next level of capability is to be able to not only describe complex discipline models, but also to describe the interactions of these models. These semantic services should provide the higher level constraints on interrelationships needed for automatically ordering the simulation components and data transformations of the various models in response to certain types of queries. For this it is necessary to represent the multi-disciplinary relationships for the many systems that make up, e.g., the Terrestrial Biogeoscience environment / “super model” described above, and the types of questions that might be asked related to these models.

This is a critical capability. As we tackle broader and more realistic problems in science, problem solving will always be the result of multi-disciplinary simulation and data analysis. On the other hand, to realize the full benefit of this process it must be available to a wide range of practitioners. If we have to assemble a team of experts representing each discipline of the multi-disciplinary model every time we want to make changes to study a slightly different phenomenon, then the utility of multi-disciplinary modeling will be very limited. We need to encode enough discipline knowledge in general semantic models so that “what if” questions can be answered in specific areas. In other words, we need to be able to have sub-discipline specialists change their model components or change their model configuration, but still be able to use the larger discipline model to ensure overall correctness of configuration while they experiment with a few subsystems or solve their specific problems, without having to consult with other experts.

Further, we would like practitioners who are not necessarily specialists in any of the disciplines of a multi-disciplinary model to be able to reliably use the mode in useful ways. A jet engine is simulated by a complex, multi-disciplinary simulation to characterize its operation, however all that an aircraft designer wants to know is how that simulation must be configured to provide the appropriate responses when coupled to a particular aircraft design, atmospheric conditions, etc.

This sort of scenario is characterized by the following two examples, where answering the what-if question requires assembling different components in different ways within a general discipline model framework that imposes constraints to ensure correct operation of the combined components. In other words, different combinations of sub-models are put together automatically.

An easily described example that illustrates the point might be given as follows. What does the itinerary look like if I wish to go San Francisco to Paris and then to Bucharest. In Bucharest, I want a 3 or 4 star hotel that is within 3 km of the Palace of the Parliament, and the hotel cost may not exceed the U. S. Dept. of State, Foreign Per Diem Rates.

To answer such a question – relatively easy, but tedious, for a human – the system must “understand” the relationships between maps and locations, between per diem charts and published hotel rates, and it must be able to apply constraints (< 3 km, 3 or 4 star, cost < $ per diem rates, etc.)

A much more realistic query of this type, related to the climate model described above, is as follows[a].

Consider a prototype query such as: “Within 20%, what will be the water runoff in the creeks of the Comanche National Grassland if we seed the clouds over southern Colorado in July and August next year.”

To answer such a question today one would have to understand the detailed descriptions of models and data for precipitation, evapotranspiration, evaporation, figure out what runoff basins are involved in the Comanche National Grassland, locate stream network models, obtain historical cloud cover data for July and August, determine the inputs and outputs for an appropriate precipitation process chain model to characterize seeding results, incorporate historical (or current) stream runoff rates, etc.

Each of these models will have certain characteristics that establish resolution, accuracy, regions of validity, etc. The data will have to be transformed into specific input types with specific units, girding, etc., in order that it may be used with the available numerical simulations.

All of this information will be contained in documentation for the models, on-line dataset descriptions, reference documents describing the accuracy of the data, etc.

A human would have to extract this information and identify appropriate data conversion programs, figure out how the models relate to each other, set up scripts to run the models and data conversions, organize the intermediate files so that downstream processing steps may refer back to them, etc.

[a] Barney Pell, Keith Golden, and Piyush Mehrotra of NASA Ames Research Center contributed to this example.

This is a process that is likely to take weeks or months in order to assemble all of the required information and gain understanding of the models needed to correctly structure all of the required operations.

On the other hand, all of the information associated with the models and data, their operational constraints, etc., can be encoded in metadata about all of the related data and about the services (component input and output data structures). Ontologies can represent the relationships among related components and data, and about the accuracy and resolution dependencies, etc.

The geographic boundaries of the Comanche National Grassland intersected with the runoff basin and stream locations will yield the hydrology basin, and the associated stream networks. Ontologies describing hydrologic simulations should give relationships among the precipitation, evapotranspiration, evaporation models, their required input, etc. Ontologies describing atmospheric moisture data from EOS satellites should provide appropriate transformations to get this data into the form required by the models.

Higher level ontologies describing relationships among the relevant geo-physical systems should describe how to establish the relationships among the sub-system level ontologies describing the models noted above. From these system level ontologies, tools should construct generalized workflows to get from cloud seeding to runoff. Similar tools applied to the sub-system models will fill out the workflow, generating abstract Grid workflows that specify data to obtain, simulations to run, intermediate files to store, etc.

This abstract / general Grid workflow is then passed to an AI based planner that constructs and adaptively manages the actual code execution and data movement. That is, the abstract, or high level, workflow describes the relationships among the simulation components and data for solving a particular problem / query. The AI planner optimizes the use of available computing and storage systems that are used to execute the simulations and manage the resulting data in a dynamic environment where computers and storage systems may come and go, may fail, etc. Moving the simulation codes and data around within the pool of available resources in order to keep the general model workflow progressing toward a solution is a problem distinct from the construction of the original, high-level workflow. See, e.g., [15,16].

Given tools that can apply constraint queries against the ontologies, the user should be able to decompose the question into a few constituent parts and fairly quickly find out how the parts must interoperate, what data is needed, and so forth. Ontologies associated with the data should describe how transformations of coordinates and units, change resolution, etc., may be accomplished, what are the required data transformations and how they must be configured.

Even assuming that the relevant data must be located manually (we are not assuming a broad data discovery capability in these examples – that is a separate topic) the end result is that the tedious human interventions are avoided, making possible a much broader use of the complex, underlying knowledge and information base, and much more easily answering the top level planning / prediction / strategy questions of non-discipline specialist users.

Conclusion

A non-specialist should be able to formulate quantitative or qualitative, declarative or constraint based queries in problem solving environments that involve multiple related data and simulations operating in several discipline models. Semantic models and tools should generate correctly structured sets of operations - sequencing and parameterizations – and also manages acquiring or generating appropriate data that is input to the analysis and simulations that will resolve the query.

This should be possible across multiple domain models, e.g. topography, hydrology, and climate illustrated in the Terrestrial Biogeoscience example above. The general data and simulation workflow must be automatically be mapped onto appropriate compute and data resources using Grid resource brokering and planning services – appropriate being determined by effective use, services that are located on specific / unique resources. All of this involves integrating and extending the integration of AI techniques and tools with Grid services technology to produce a Semantic Grid.

We recognize that the first two categories of semantic services given above are probably within the scope of current technology, and that the third is more visionary. However, these are the sorts of services that are needed to move Grids to a central position in the next generation science process.

Acknowledgements

The author would like to thank Dieter Fensel and Mark Musen as guest editors of the March/April, 2001 issue of IEEE Intelligent Systems, together with the authors who contributed to that issue on Web ontology languages, for a solid introduction to this topic that will be at the heart of the semantic services envisioned here. Thanks also to Horst Simon of the LBNL Computational Research Division for comments on the first draft of this paper, and to an anonymous reviewer whose comments resulted in a number of clarifications.

This work was funded by the U.S. Dept. of Energy, Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Sciences Division, Collaboratory Environments Program under contract DE-AC03-76SF00098 with the University of California, and by NASA's Aero-Space Enterprise, Computing, Information, and Communication Technologies (CICT) Program, Computing, Networking, and Information Systems Project. This document is LBNL report number 54249.

5 References

[1] *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, eds. 2nd. ed. 2003, Morgan Kaufmann.

[2] *Grid Computing: Making the Global Infrastructure a Reality*, F. Berman, G. Fox and T. Hey, eds. 2nd. ed. 2003, John Wiley. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470853190.html

[3] Chimera: A Virtual Data System for Representing, Querying, and Automating Data Derivation, I. Foster, J. Voeckler, M. Wilde and Y. Zhao. In *14th Intl. Conf. on Scientific and Statistical Database Management.* 2002. Edinburgh, Scotland.http://www.griphyn.org/documents/document_server/show_docs.php?category=tech&id=156

[4] XCAT 2.0: A Component-Based Programming Model for Grid Web Services, M. Govindaraju, S. Krishnan, K. Chiu, A. Slominski, D. Gannon and R. Bramley. June, 2002, Indiana University, Bloomington, Indiana. http://www.extreme.indiana.edu/xcat/

[5] Middleware Technology to Support Science Portals: a Gateway to the Grid, D. Gannon and R. Bramley. U. S. Dept. of Energy, National Collaboratories Program. http://doecollaboratory.pnl.gov/research2/

[6] iVDGL - International Virtual Data Grid Laboratory. http://www.ivdgl.org/

[7] NEESgrid - Network for Earthquake Engineering Simulation Grid. http://www.neesgrid.org/

[8] Virtual Observatories of the Future. http://www.astro.caltech.edu/nvoconf/

[9] GGF - Global Grid Forum. http://www.gridforum.org/

[10] High Performance Network Planning Workshop, DOE Office of Science. August 13-15, 2002: Reston, Virginia, USA. http://doecollaboratory.pnl.gov/meetings/hpnpw

[11] DOE Science Networking Challenge: Roadmap to 2008, DOE Office of Science. June 3-5, 2003: Reston, Virginia, USA. http://gate.hep.anl.gov/lprice/Roadmap/index.html

[12] Blueprint for Future Science Middleware and Grid Research and Infrastructure, LSN-MAGIC. August, 26-28, 2002: Chicago, Illinois, USA. http://www.nsf-middleware.org/MAGIC/

[13] A Science Based Case for Large Scale Simulation, DOE Office of Science. June 24-25, 2003: Arlington, Virginia, USA. http://www.pnl.gov/scales/

[14] ESG - Earth System Grid. http://www.earthsystemgrid.org/

[15] Mapping Abstract Complex Workflows onto Grid Environments, E. Deelman, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, K. Vahi, A. Lazzarini, A. Arbree, R. Cavanaugh and S. Koranda. Journal of Grid Computing, 2003. 1(1): p. 9-23,. http://www.isi.edu/~blythe/papers/grid-journal03.html

[16] A Case Study: Using Workflow And Ai Planners, M. D. R-Moreno, P. Kearney and D. Meziat. In *The 19th Workshop of the UK Planning and Scheduling Special Interest Group (PLANSIG 2000)*. 2000. The Open University, Milton Keynes (England).http://mcs.open.ac.uk/plansig2000/Papers/MD.R-Moreno.pdf